Few-shot Learning with Multilingual Generative Language Models

Abstract

Large-scale generative language models such as GPT-3 are competitive few-shot learners. While these models are known to be able to jointly represent many different languages, their training data is dominated by English, potentially limiting their cross-lingual generalization. In this work, we train multilingual generative language models on a corpus covering a diverse set of languages, and study their few- and zero-shot learning capabilities in a wide range of tasks. Our largest model with 7.5 billion parameters sets new state of the art in few-shot learning in more than 20 representative languages, outperforming GPT-3 of comparable size in multilingual commonsense reasoning (with +7.4% absolute accuracy improvement in 0-shot settings and +9.4% in 4-shot settings) and natural language inference (+5.4% in each of 0-shot and 4-shot settings). On the FLORES-101 machine translation benchmark, our model outperforms GPT-3 on 171 out of 182 directions with 32 training examples, while surpassing the official supervised baseline in 45 directions. We conduct an in-depth analysis of different multilingual prompting approaches, showing in particular that strong few-shot learning performance across languages can be achieved via cross-lingual transfer through both templates and demonstration examples.